Thanks to Brandi Loper

a.
```
data whas;
filename ps5 url "http://algebra.sci.csueastbay.edu/~sfan/SubPages/CSUteach/st6652/homework/ps5.dat";
infile ps5 firstobs=2;
input id age gender cpk chf miord time stat;
run;
```

```
proc phreg data=whas;
  model time*stat(0)=age gender cpk chf miord;
run;
```

```
proc phreg data=whas;
  model time*stat(0)=age gender cpk chf miord/ties=efron;
run;
```

```
proc phreg data=whas;
  model time*stat(0)=age gender cpk chf miord/ties=exact;
run;
```

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Breslow</th>
<th></th>
<th>Efron</th>
<th></th>
<th>Exact</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Estimate</td>
<td>S.E.</td>
<td>Estimate</td>
<td>S.E.</td>
<td>Estimate</td>
</tr>
<tr>
<td>age</td>
<td>0.03551</td>
<td>0.00592</td>
<td>0.03452</td>
<td>0.00592</td>
<td>0.03452</td>
</tr>
<tr>
<td>gender</td>
<td>0.08301</td>
<td>0.13349</td>
<td>0.08359</td>
<td>0.13349</td>
<td>0.08359</td>
</tr>
<tr>
<td>cpk</td>
<td>0.0001498</td>
<td>0.0000617</td>
<td>0.0001532</td>
<td>0.0000617</td>
<td>0.0001533</td>
</tr>
<tr>
<td>chf</td>
<td>0.76790</td>
<td>0.13615</td>
<td>0.77039</td>
<td>0.13612</td>
<td>0.77040</td>
</tr>
<tr>
<td>miord</td>
<td>0.39847</td>
<td>0.12966</td>
<td>0.40068</td>
<td>0.12966</td>
<td>0.40070</td>
</tr>
</tbody>
</table>

Nearly all the estimates and standard errors are the same with some differences; however, Efron’s approximation gave results that were much closer to the exact results than Breslow’s approximation. When dealing with a lot of ties we would thus consider using the exact method, which would be the most accurate, but in this case, since Efron’s approximation is adequate to use and is less computationally demanding, I recommend this method. You can choose Exact too if your computer is sufficiently powerful.

b.
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
```
**proc phreg** data=whas;
  model time*stat(0)=age cpk chf miord/ties=efron;
run;

<table>
<thead>
<tr>
<th>-2 Log L (full model)</th>
<th>-2 Log L (reduced model)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2720.987</td>
<td>2721.378</td>
</tr>
</tbody>
</table>

**LR test:**

$H_0: \beta_{gender} = 0$ (reduced model)

$H_1: \beta_{gender} \neq 0$ (full model)

$\alpha = .05$ significance level

**Test statistic:**

$\Lambda = -2[\text{log L (reduced model)} - \text{log L (full model)}] = 2721.378 - 2720.987 = 0.391 \sim \chi^2$ with df = # of parameters in full model - # of parameters in reduced model = 5-4 = 1,

$p$-value = 1-pchisq(0.391,1) = 0.5317742

**Decision:**

Reject $H_0$ if $\Lambda > \chi^2(df=1) = 3.84$, conclude there is a difference in the two models (full)

Do not reject $H_0$ if $\Lambda < \chi^2(df=1) = 3.84$, conclude there is not a difference in the two models (reduced)

**Conclusion:**

Since $\Lambda = 0.391 < \chi^2(df=1) = 3.84$, we do not reject $H_0$ and conclude that there is no difference between the two models, so we will then choose the reduced model.

c.

**proc phreg** data=whas;
  model time*stat(0)=age gender cpk chf miord/ties=efron
  selection=backward;
run;

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
<th>Hazard Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>1</td>
<td>0.03640</td>
<td>0.00571</td>
<td>40.6439</td>
<td>&lt;0.001</td>
<td>1.037</td>
</tr>
<tr>
<td>cpk</td>
<td>1</td>
<td>0.0001508</td>
<td>0.0000617</td>
<td>5.9743</td>
<td>0.0145</td>
<td>1.000</td>
</tr>
<tr>
<td>chf</td>
<td>1</td>
<td>0.77150</td>
<td>0.13600</td>
<td>32.1824</td>
<td>&lt;0.001</td>
<td>2.163</td>
</tr>
<tr>
<td>miord</td>
<td>1</td>
<td>0.40290</td>
<td>0.12962</td>
<td>9.6615</td>
<td>0.0019</td>
<td>1.496</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Step</th>
<th>Effect Removed</th>
<th>DF</th>
<th>Number In</th>
<th>Wald</th>
<th>Pr &gt; ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>gender</td>
<td>1</td>
<td>4</td>
<td>0.3921</td>
<td>0.5312</td>
</tr>
</tbody>
</table>
Here we can see that the backward model selection shows our final model is the same as our reduced model. So the L-R test for the adequacy of our final model is as before.

d.  

```plaintext
proc phreg data=whas;
   model time*stat(0)=age cpk chf miord/ties=efron;
run
```

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>Chi-Square</th>
<th>Pr &gt; ChiSq</th>
<th>Hazard Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>age</td>
<td>1</td>
<td>0.03640</td>
<td>0.00571</td>
<td>40.6439</td>
<td>&lt;.0001</td>
<td>1.037</td>
</tr>
<tr>
<td>cpk</td>
<td>1</td>
<td>0.0001508</td>
<td>0.0000617</td>
<td>5.9743</td>
<td>0.0145</td>
<td>1.000</td>
</tr>
<tr>
<td>chf</td>
<td>1</td>
<td>0.77150</td>
<td>0.13600</td>
<td>32.1824</td>
<td>&lt;.0001</td>
<td>2.163</td>
</tr>
<tr>
<td>miord</td>
<td>1</td>
<td>0.40290</td>
<td>0.12962</td>
<td>9.6615</td>
<td>0.0019</td>
<td>1.496</td>
</tr>
</tbody>
</table>

The fitted model is:

\[
h_x(y) = h_0(y)^*\exp\{\beta^T x\}
\]

\[
= h_0(y)^*\exp\{0.0364*age + 0.0001508*cpk + 0.77150*chf + 0.40290*miord\}
\]

/* Baseline cumulative hazard functions */

data b;
   set a;
   negls=-ls;
run;
proc gplot data=b;
   title "Baseline Cumulative Hazard Function";
   plot negls*time;
run;

Interpretation of the effect of each covariate:

Gender is an insignificant covariate.

While holding all other covariates constant, for the effect of a 1 unit change in age, the relative risk is \( \exp(0.0364) = 1.037 \), which means as a patient’s age increases by 1 year the risk of death increases by 3.7%.

While holding all other covariates constant, for the effect of a 100 unit change in cpk, the relative risk for cpk is \( \exp(0.0001508 \times 100) = 1.015 \), which means as a patient’s cpk increases by 100 international units the risk of death increases by 1.5%.

While holding all other covariates constant, the risk of death for a patient with left heart failure complications is \( \exp(0.77150) = 2.163 \) times the risk (or 116% higher) for a patient without no such complications.

While holding all other covariates constant, the risk of death for a patient with a recurrent heart is \( \exp(0.4029) = 1.496 \) times the risk (or 49.6% higher) for a patient with first heart attack.

e.

```plaintext
/* predicted survival function */
data x;
input age cpk chf ord;
cards;
62 485 0 0
62 485 1 0
run;

proc phreg data=whas;
  model time*stat(0)=age cpk chf ord
    / ties = efron;
  baseline out=b covariates=x survival=s;
```
run;

proc print data=b;
run;

goptions reset=global;
legend1 label=("Left heart failure complications") value=("No" "Yes") position=(top left inside);

proc gplot data=b;
title "Predicted Survival Function";
symbol1 value=dot color=black i=join;
symbol2 value=triangle color=black i=join;
plot s*time=chf/legend=legend1;
run;

For a patient at age 62, having cpk 485, no recurrence and has left failure complications
the survival curve falls a lot faster than a patient who doesn’t have any heart
complications, this means the survival rate for a patient with left heart failure
complications is a lot lower than a patient with no such complications.